Communication Theory II

Lecture 5: Review on Probability Theory

Ahmed Elnakib, PhD
Assistant Professor, Mansoura University, Egypt

Febraury 22th, 2015

Lecture Outlines

- Review on probability theory
  - Conditional probability and Bayes' rule
  - Random variables and the distribution functions
  - Multiple random variables
  - The concept of expectation and higher order statistics

Conditional probability

- Probability of event *A* given that event *B* has occurred

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

- If *A, B* are independent: $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$
  - At knowledge of the occurrence of either event, *A* or *B*, tells us no more about the probability of occurrence of the other event than we knew without that knowledge
  - Events A and B that satisfy this condition are said to be independent, and then $\mathbb{P}[A|B] = \mathbb{P}[A]$

Bayes' Rule

- Posterior probability: inference based on make assumptions

$$\mathbb{P}[A \cap B] = \mathbb{P}[A|B]\mathbb{P}[B]$$

$$\mathbb{P}[A \cap B] = \mathbb{P}[B|A]\mathbb{P}[A]$$

$$\mathbb{P}[A|B]\mathbb{P}[B] = \mathbb{P}[B|A]\mathbb{P}[A]$$

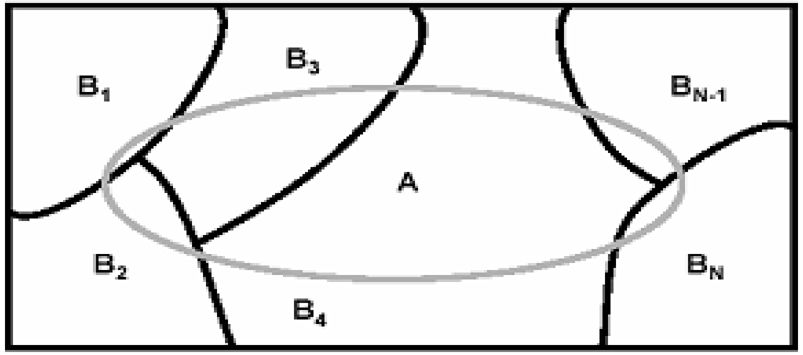
$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\mathbb{P}[B]}{\mathbb{P}[A]}$$

Law of Total Probability

Let $B_1, \ldots B_N$ be N mutually exclusive events*, whose union gives the sample space Ω. Hence the events B constitute a partition of Ω

Now consider an event A, a subset of Ω. This event can be represented as

$$A = A \cap \Omega = A \cap (B_1 \cup B_2 \cup \ldots \cup B_N) = (A \cap B_1) \cup (A \cap B_2) \cup \ldots (A \cap B_N)$$



Since the B_i are mutually exclusive

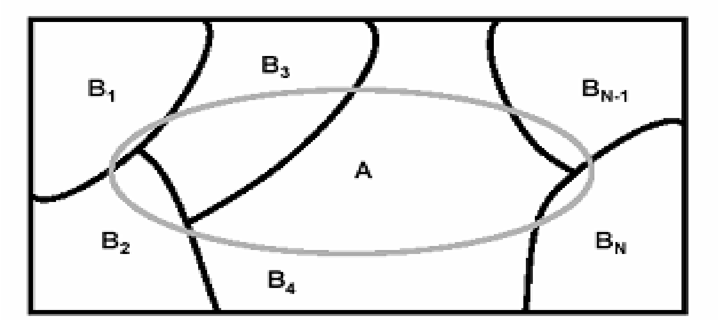
$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_N)$$

$$P(A) = P(A \mid B_1)P(B_1) + \ldots + P(A \mid B_N)P(B_N) = \sum_{k=1}^{N} P(A \mid B_k)P(B_k)$$

* Two events *A, B* are **mutually exclusive** if they cannot occur at the same time: $A \cap B = \emptyset$

General Bayes Rule

- We now pose the following question: Given that the event A has occurred. What is the probability that any single one of the event B's occur?



Rev. Thomas Bayes,
(1702-1761)

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A \mid B_j) \cdot P(B_j)}{\sum_{k=1}^{N} P(A \mid B_k) \cdot P(B_k)}$$

This is known as the Bayes rule

Example

Radar, a remote sensing system, operates by transmitting a sequence of pulses and has its receiver listen to echoes produced by a target (e.g., aircraft) that could be present in its surveillance area.

Let the events A and B be defined as follows:

$A = \{\text{a target is present in the area under surveillance}\}$
$A^c = \{\text{there is no target in the area}\}$
$B = \{\text{the radar receiver detects a target}\}$

In the radar detection problem, there are three probabilities of particular interest:

$\mathbb{P}[A]$	probability that a target is present in the area; this probability is called the *prior probability.*
$\mathbb{P}[B\|A]$	probability that the radar receiver detects a target, given that a target is actually present in the area; this second probability is called the *probability of detection*.
$\mathbb{P}[B\|A^c]$	probability that the radar receiver detects a target in the area, given that there is no target in the surveillance area; this third probability is called the *probability of false alarm.*

Given:

$$\mathbb{P}[A] = 0.02$$
$$\mathbb{P}[B|A] = 0.99$$
$$\mathbb{P}[B|A^c] = 0.01$$

Required: $\boxed{\mathbb{P}[A|B]}$

The probability that a target is present in the surveillance area given that the radar receiver has made a target detection.

solution

Posterior probability

likelihood

Prior probability

evidence

Provide the language for inference

$$\mathbb{P}[A] = 0.02$$
$$\mathbb{P}[B|A] = 0.99$$
$$\mathbb{P}[B|A^c] = 0.01$$

$$\begin{aligned}
\mathbb{P}[A|B] &= \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]} \\
&= \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|A^c]\mathbb{P}[A^c]} \\
&= \frac{0.99 \times 0.02}{0.99 \times 0.02 + 0.01 \times 0.98} \\
&= \frac{0.0198}{0.0296} \\
&\approx 0.69
\end{aligned}$$

Random Variables

- Real experiments involve using one or more real valued quantities called random variables
- The outcomes are associated with some numerical values of interest (random variable)

The **random variable** is a function whose domain is a sample space and whose range is some set of real numbers

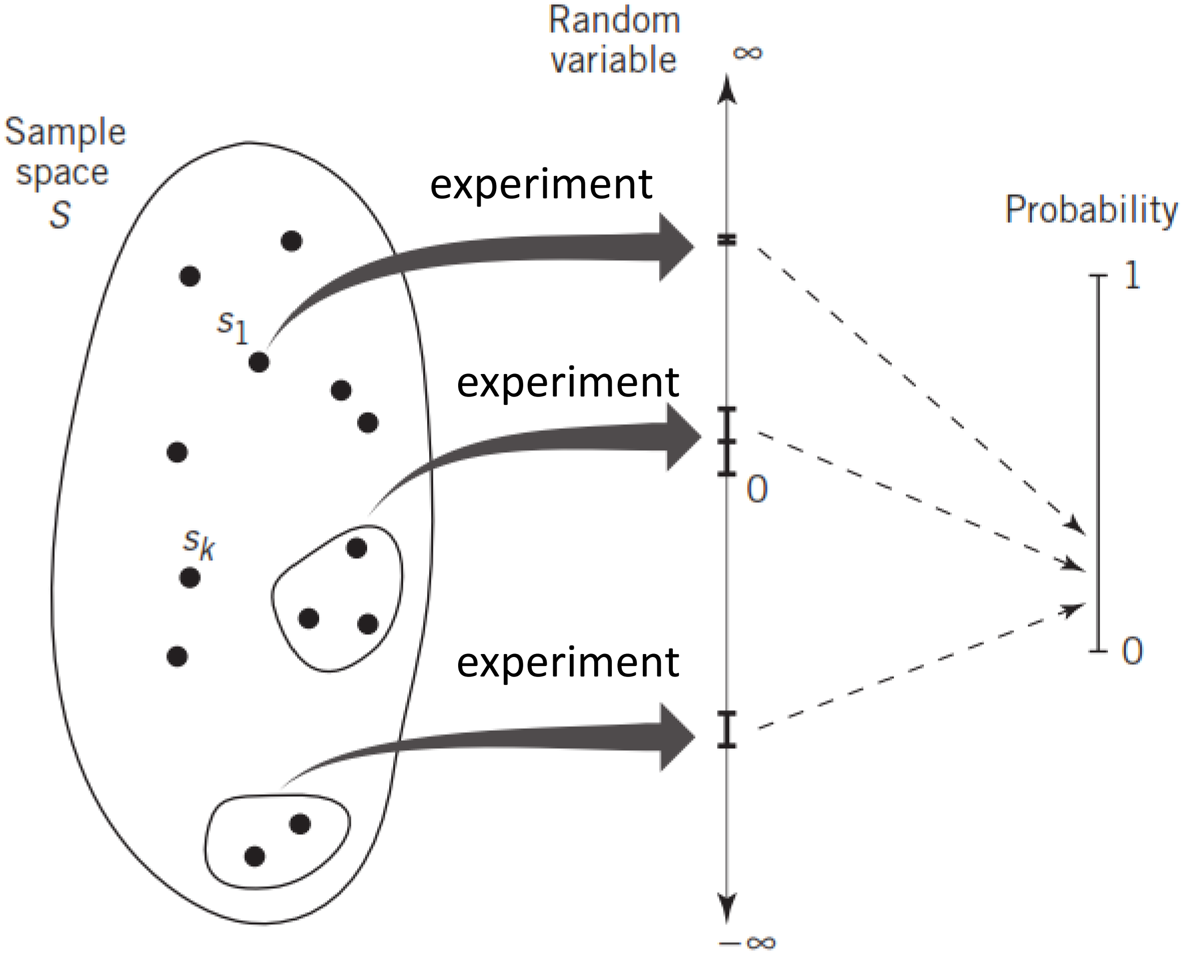


Illustration of the relationship between sample space, random variables, and probability

Discrete and Continuous Random Variables

- Random variables can **discrete**, e.g., the number of heads in three consecutive coin tosses, or **continuous**, the weight of a class member.

Probability/Cumulative Mass Function

- A **probability mass (distribution) function,** $p(X=x)$, is a function that tells us the probability of x, an observation of X, assuming a specific value

- The **cumulative mass (distribution) function,** $F_X(x)$, indicates the probability of X assuming a value less then or equal to x

$$F_X(x) = \mathbb{P}[X \leq x] \qquad \text{for all } x$$

  - Lies between zero and one: $F_X(x=-\infty)=0$ and $F_X(x=\infty)=1$
  - Monotone non-decreasing: $F_X(x_1) \leq F_X(x_2)$ for $x_1 < x_2$

Probability/Cumulative Density Function (cont’d)

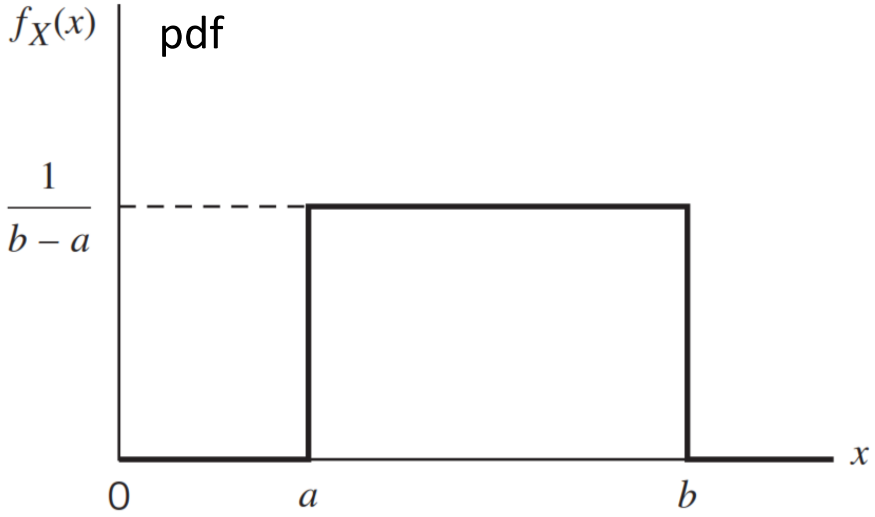
- For continuos variables:
  - Probability Mass Function → Probability Density function (pdf),
    - $f(x) = \frac{d}{dx} F_X(x)$
    - Non-negativity: $f_X(x) \geq 0$
    - Normalization (total area under the curve=1), i.e., $\int_{-\infty}^{\infty} f_X(x)dx = 1$

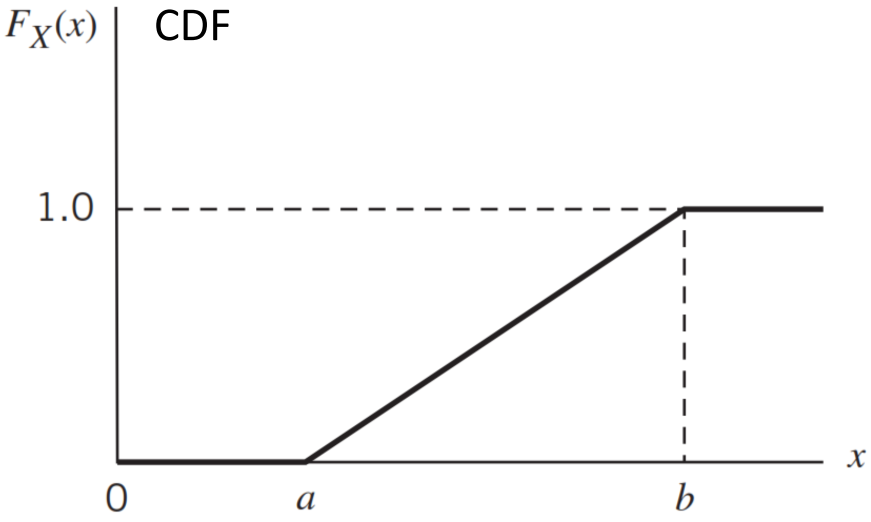
  - Cumulative Mass Function → Cumulative Density Function (CDF)
    - $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx$
    - $F_X(x_1) = \int_{-\infty}^{x_1} f_X(x)dx$

Example 1 Uniform Distribution

$$f_X(x) = \begin{cases} 0, & x \le a \\ \dfrac{1}{b-a}, & a < x \le b \\ 0, & x > b \end{cases}$$

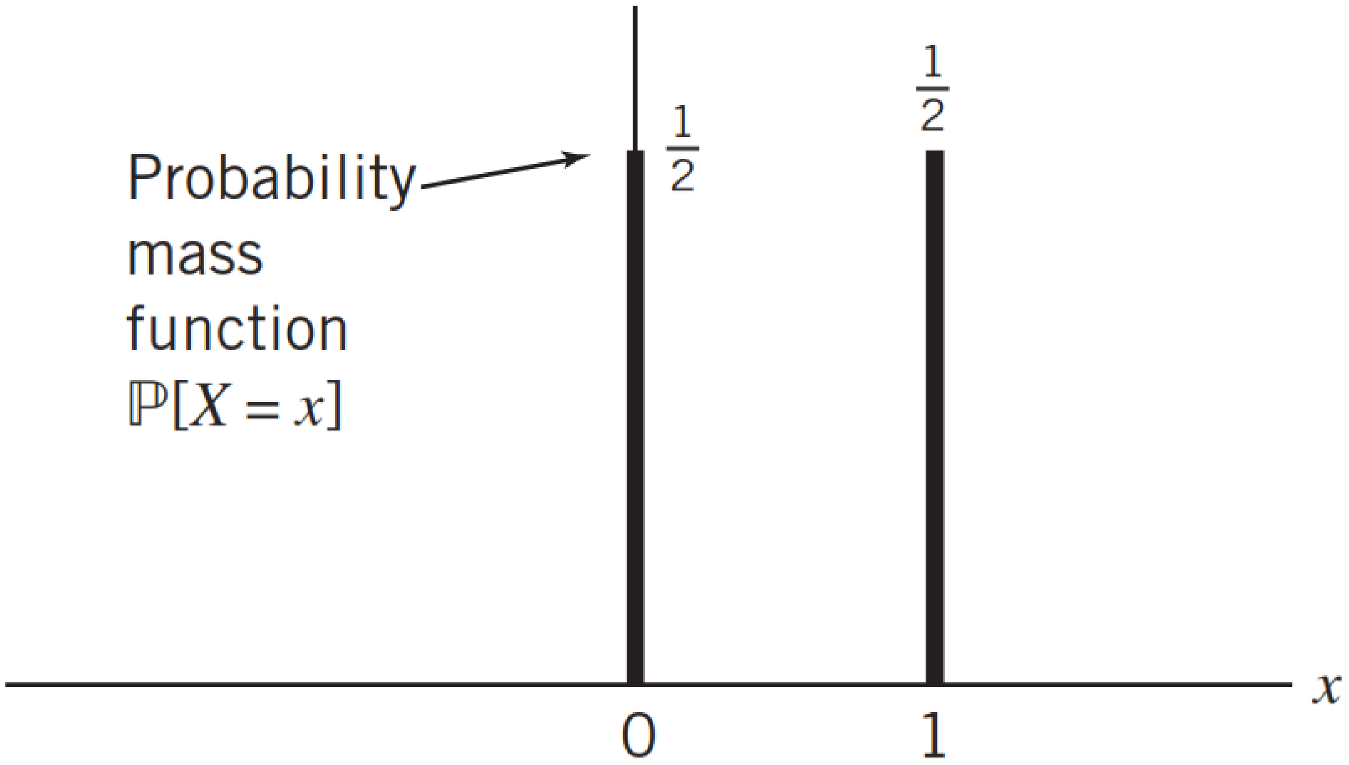
$$F_X(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ 1 & x > b \end{cases}$$





Example 2 Bernoulli Random Variable

$$p_X(x) = \begin{cases} 1-p & x = 0 \\ p, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$



Illustrating the probability mass function for a fair coin-tossing experiment

Multiple Random Variables

- The **joint distribution function** $F_{X,Y}(x,y)$ is the probability that the random variable X is less than or equal to a specified value x, and that the random variable Y is less than or equal to another specified value y

$$F_{X,Y}(x,y) = \mathbb{P}[X \leq x, Y \leq y] = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

- The **joint probability density function** $f_{X,Y}(x,y)$

For continuous R.V.

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

Total volume under the joint probability density graph

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$$

The joint probability density function

- The joint probability density function $f_{X,\,Y}(x, y)$ contains all the conceivable information on the two continuous random variables X *and* Y that is needed for the probability analysis of joint random variables

Marginal probability

$$F_X(x) = F_{X,Y}(x, \infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

Marginal probability distribution

$$F_Y(y) = F_{X,Y}(\infty, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}F_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y$$

Marginal probability density

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x$$

Conditional Probability Density Function

- The conditional probability density function of Y, such that $X = x$

$$f_{Y|X}(y|x) = f_Y(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

- By cross multiplication:

$$\underbrace{f_{X,Y}(x, y)}_{\text{Joint probability}} = \underbrace{f_Y(y|x)}_{\text{conditional probability}}\underbrace{f_X(x)}_{\text{Marginal probability}}$$

Axioms of probability

$$f_Y(y|x) \geq 0$$

$$\int_{-\infty}^{\infty} f_Y(y|x)\mathrm{d}y = 1$$

- Y, X are independent if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$

  and then $f_Y(y|x) = f_Y(y)$

Expectation: first order statistical average

- The probability mas function of a random variable X provides us with several numbers, the probabilities of all the possible values of X
- It would be desirable to summarize this information in a single representative number
- This is accomplished by the **expectation** of X, which is a weighted (in proportion to probabilities) average of the possible values of X
- **Expectation** of X Locates the center of gravity of the area under the probability density curve of the random variable X

$$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$$

For continuous R.V.

$$\mathbb{E}[X] = \sum_x x p_X(x)$$

For discrete R.V.

Properties of Expectation

- **Linearity:** if $Z = X + Y \rightarrow \mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y]$

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i])$$

- **Statistical independence:** if X and Y are independent R.Vs then:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

- Note that: $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, \mathrm{d}x$

Higher Order Statistics

- n_{th} moment of the probability distribution of a random variable X $\mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,\mathrm{d}x$
- Mean square value of X $\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,\mathrm{d}x$
- Central moments $\mathbb{E}[(X-\mu_X)^n] = \int_{-\infty}^{\infty} (x-\mu_X)^n f_X(x)\,\mathrm{d}x$
- Variance $\mathrm{var}[X] = \mathbb{E}(X-\mu_x)^2 = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\,\mathrm{d}x = \mathbb{E}[X^2]-\mu_X^2$
- Chebyshev inequality $\mathbb{P}[|X-\mu_X| \geq \varepsilon] \leq \dfrac{\sigma_X}{\varepsilon^2}$

Moments of Multiple Random Variables

- The joint moments (expectations of $X^i\ Y^j$ for positive i,j) $\mathbb{E}[X^i Y^k] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^i y^k f_{X,\,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y$
- Correlation (if $i=j=1$) $\mathbb{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\ y\ f_{X,\,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y$
- Covariance of X *and* Y *is* the correlation of the centered random variables $(X - \mathrm{E}[X])$ and $(Y - \mathrm{E}[Y])$

$$\mathrm{cov}[XY] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mu_X\mu_Y$$

- Correlation coefficient of X and Y $\rho(X, Y) = \dfrac{\mathrm{cov}[XY]}{\sigma_X\sigma_Y}$
- Covariance of uncorrelated R.Vs X and Y $\mathrm{cov}[XY] = 0$

$\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$

σ_X^2 and σ_Y^2 are variances of X and Y

Uncorrelated vs. Orthogonal R.Vs

The two random variables X and Y are said to be *uncorrelated* if, and only if, their covariance is zero; that is,

$$\operatorname{cov}[XY] = 0$$

They are said to be *orthogonal* if and only if their correlation is zero; that is,

$$\mathbb{E}[XY] = 0$$

If one of the random variables X *and* Y or both have zero means, and if they are orthogonal, then they are uncorrelated, and vice versa

$$\operatorname{cov}[XY] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mu_X\mu_Y$$

Questions